Hotel Booking Cancellation Prediction

Contents / Agenda

- Data Dictionary

- Business Problem Overview and Solution Approach

- EDA Results

- Data Preprocessing

- Model Performance Summary

- Conclusions and Recommendations

Data Dictionary

The data contains the different attributes of **customers' booking details**. The detailed data dictionary is given below:

- **Booking_ID:** the unique identifier of each booking
- **no_of_adults:** Number of adults
- **no_of_children:** Number of Children
- **no_of_weekend_nights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **no_of_week_nights:** Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **required_car_parking_space:** Does the customer require a car parking space? (0 - No, 1- Yes)
- **room_type_reserved:** Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels Group

Data Dictionary

- **lead_time:** Number of days between the date of booking and the arrival date

- **arrival_year:** Year of arrival date

- **arrival_month:** Month of arrival date

- **arrival_date:** Date of the month

- **market_segment_type:** Market segment designation.

- **repeated_guest:** Is the customer a repeated guest? (0 - No, 1- Yes)

- **no_of_previous_cancellations:** Number of previous bookings that were canceled by the customer prior to the current booking

- **no_of_previous_bookings_not_canceled:** Number of previous bookings not canceled by the customer prior to the current booking

How to use this deck?

- This slide deck serves as a comprehensive template for your case study.

- Within this deck, you will come across various questions that are intended to test your ability to understand data visualizations, discover patterns / insights and postulate hypothesis. Think thoroughly and provide answers to these questions

- Please feel free to incorporate additional points if you deem necessary

Business Problem Overview

A significant number of hotel bookings are called off due to cancellations or no-shows. The typical reasons for cancellations include change of plans, scheduling conflicts, etc. This is often made easier by the option to do so free of charge or preferably at a low cost which is beneficial to hotel guests but it is a less desirable and possibly revenue-diminishing factor for hotels to deal with. Such losses are particularly high on last-minute cancellations.

The new technologies involving online booking channels have dramatically changed customers' booking possibilities and behavior. This adds a further dimension to the challenge of how hotels handle cancellations, which are no longer limited to traditional booking and guest characteristics.

The cancellation of bookings impact a hotel on various fronts:

1. Loss of resources (revenue) when the hotel cannot resell the room.
2. Additional costs of distribution channels by increasing commissions or paying for publicity to help sell these rooms.
3. Lowering prices last minute, so the hotel can resell a room, resulting in reducing the profit margin.
4. Human resources to make arrangements for the guests.

Objective

- The increasing number of cancellations calls for a Machine Learning based solution that can help in predicting which booking is likely to be canceled. INN Hotels Group has a chain of hotels in Portugal, they are facing problems with the high number of booking cancellations and have reached out to your firm for data-driven solutions. You as a data scientist have to analyze the data provided to find which factors have a high influence on booking cancellations, build a predictive model that can predict which booking is going to be canceled in advance, and help in formulating profitable policies for cancellations and refunds.

Expectation :

1 - build a model using any tool you perfer. You will be requested to show the model and the outcome from the model in the final interview.

2 - You need to show charts as per the need. You need to attach the charts in this document. (each slide has a question and you might need to answer this question based on some charts.

3 – full EDA is required.

4 - You might need to pre-process the dataset.

Business Problem Overview and Solution Approach

- Please define the problem

- Please mention the solution approach / methodology

EDA - Univariate Analysis

- Explore strategies to effectively manage and accommodate the increased demand

EDA - Univariate Analysis

- Identify the percentage of bookings made online versus offline and highlight any trends or insights that can inform business strategies.

EDA - Univariate Analysis

- Analyze the percentage of bookings that are cancelled. Are the cancellation trends a serious issue that needs to be looked at? If so, what potential solutions can be implemented to address this problem?

EDA - Univariate Analysis

- To what extent are customers returning to the hotel? Is this indication considered positive or negative, and what factors may be influencing this behavior? Furthermore, what strategies can be implemented to address and potentially improve this situation?

EDA - Univariate Analysis

- Explore how hotels can optimize their revenue management strategies considering the significant price variations observed across different market segments

Results EDA - Univariate Analysis

- Provide a concise overview of the discoveries derived from the Univariate Analysis.

EDA - Bivariate Analysis

- Explore strategies to mitigate cancellations and consider how these insights can be leveraged to reduce cancellations across different booking channels

EDA - Bivariate Analysis

- Examine the influence of special requests on booking cancellations and devise effective strategies to mitigate cancellation rates when customers make such requests.

EDA - Bivariate Analysis

- What insights can be gained from the lead time of canceled bookings, and how can we leverage this information to improve our booking system and enhance customer satisfaction?

EDA - Bivariate Analysis

- Analyze the cancellation rate for each month and compare it to other months to determine any patterns or reasons for variation.

Results EDA - Bivariate Analysis

- Provide a concise overview of the discoveries derived from the Bivariate Analysis.

Model Performance Evaluation Decision Tree

- Please comment on the model performance of the Decision Tree Model

Model	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision
Decision Tree						

Model Performance Evaluation Decision Tree

- To what extent and in what ways have the attribute weights influenced the overall performance of the model?

Model Performance Evaluation Pruned Decision Tree

- Based on the evaluation metrics obtained from a pruned decision tree, do you perceive any notable enhancements in performance? What factors do you believe contributed to this improved performance compared to the previous version?

Model	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision
Decision Tree						
Decision Tree - Pruned						

Model Performance Evaluation Random Forest

Model	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision
Random Forest						

Model Performance Evaluation Random Forest

- To what extent and in what ways have the attribute weights influenced the overall performance of the model?

Model Performance Evaluation Pruned Random Forest

- Based on the evaluation metrics obtained from a pruned Random Forest, do you perceive any notable enhancements in performance? What factors do you believe contributed to this improved performance compared to the previous version?

Model	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision
Random Forest						
Random Forest - Pruned						

Model Performance Evaluation Pruned Random Forest

- To what extent and in what ways have the attribute weights influenced the overall performance of the model?

Model Performance Summary

- Description of the ML model that best fits the business objective. Also state which evaluation metric (such as accuracy, recall or precision) is important to achieve the business objective and why?

- Synopsis of the key features employed by the ML model to make predictions.

- Offer suggestions and advice for the hotel based on the gathered insights.

Model	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision
Decision Tree						
Decision Tree - Pruned						
Random Forest						
Random Forest - Pruned						

Conclusions and Recommendations

- Please mention actionable insights & recommendations

- Below are some useful pointers that can guide you :

  - Identify the hotel's target market and booking patterns

  - Analyze the booking trends and cancellations over a period of time

  - Determine the impact of cancellations on hotel revenue and resources

  - Evaluate the consistency of cancellation policies across market segments

  - Determine how to improve the experience of repeated customers and attract new ones